

FACTORS ASSOCIATED WITH THE PREDICTION OF ATTENDANCE TO THE UK BREAST SCREENING PROGRAMME

H.E. Arochena¹, R.N.G. Naguib¹, H. Mashhoudy¹, M.E. Wheaton², M.G. Wallis²

¹ BIOCORE, School of Mathematical and Information Science, Coventry University, Coventry, UK

² Warwickshire, Solihull and Coventry Breast Screening Unit, Coventry and Warwickshire Hospital, Coventry, UK

Abstract - An analysis of the predictors of attendance and screening variation to the UK Breast Screening Programme over four episodes is carried out. Variables such as Townsend deprivation score and postal area (zip code) are taken into account.

Keywords - Cancer, screening variation, Townsend deprivation score, round length

I. INTRODUCTION

As part of the UK National Health Service (NHS) Breast Screening Programme, all non-symptomatic women aged 50 to 64 years old in the country are invited every three years for screening (other ages are also screened on request). This full population programme aims to reduce mortality from breast cancer in the screened population.

Essential factors to take into account in order to achieve this aim are a high attendance of the women in response to the screening invitation and a low range of screening variation if any.

The present study investigates the factors affecting attendance and screening variation in the screening programme as well as their possible predictors.

This work is part of a major research project based on data collected during the first 10 years of screening at the Warwickshire, Solihull and Coventry Breast Screening Unit. Precedents complementing this work can be found in [1-3].

II. DATA AND METHODOLOGY

Data used covers 10 years of running of the preventive breast screening programme (four screening episodes). A total of 281,415 invitations have been assessed, involving 137,051 non-symptomatic women.

The Townsend deprivation score, amongst others, is used as a measure of socio-economic deprivation. It measures the level of deprivation based on housing, income, car ownership

and educational level amongst other socio-economic factors [4].

Variables analysed include the age band of women at invitation, postal area where they live, Townsend deprivation score for that postal area and their attendance to invitation. Also taken into account are their screening variation, (difference in days between the date of invitation and actual date of screening); the screening end code resulting from the screening episode and the round length of the episode (difference in years between the date of last screening and the date of next offered appointment), amongst others.

Table I gives the nomenclature used for the coding of the screening end code (SEC).

In order to establish associations and possible predictors for categorical variables, coefficients such as Lambda (λ), Uncertainty, Phi (Φ), Cramer's V and Contingency have been measured. Those coefficients are interpreted in a similar way as the correlation coefficient. A comprehensive discussion can be found in [5-7].

In this analysis, an association (denoted by x) is classified as follows:

$0.00 \leq x \leq 0.09$	virtually no association
$0.10 \leq x \leq 0.19$	very small association
$0.20 \leq x \leq 0.39$	small association
$0.40 \leq x \leq 0.49$	medium association
$0.50 \leq x \leq 0.69$	high association
$0.70 \leq x \leq 1.00$	very high association

Only those with an association of 0.2 or higher have been considered for further analysis. These are shown in Table II.

For those pairs of variables with a high enough association coefficient value, contingency tables and their proportions based on the marginal totals have been used to establish the possible prediction of one variable, given a value of the other.

TABLE I
SCREENING END CODE (SEC) DEFINITIONS

Classification	Screening (S)	Assessment (A)	Cytology (C)	Core biopsy (W)	Histology (H)
Normal outcome (-)	S-	A-	C-	W-	H-
Abnormal outcome (abn)	Sabn	Aabn	Cabn	Wabn	Habn
Cancer Positive outcome (+)	S+	A+	C+	W+	H+

Report Documentation Page

Report Date 25 Oct 2001	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle Factors Associated With The Prediction of Attendance to The UK Breast Screening Programme		Contract Number
		Grant Number
		Program Element Number
Author(s)		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) BIOCORE, School of Mathematical and Information Science Coventry University United Kingdom		Performing Organization Report Number
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500		Sponsor/Monitor's Acronym(s)
		Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-26, 2001 held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.		
Abstract		
Subject Terms		
Report Classification unclassified		Classification of this page unclassified
Classification of Abstract unclassified		Limitation of Abstract UU
Number of Pages 4		

TABLE II
RELEVANT ASSOCIATIONS BETWEEN PAIRS OF VARIABLES

Pairs of variables	Coefficient				
	λ	Uncertainty	Φ	Cramers' V	Contingency
Ageband2 vs. Attendance2	0.75	0.68	1.00	0.70	0.70
Ageband3 vs. Attendance3	0.82	0.80	1.00	0.71	0.71
Ageband4 vs. Attendance4	0.79	0.86	1.00	0.71	0.71
Ageband2 vs. Screening End Code 2	0.73	0.49	0.78	0.35	0.61
Ageband3 vs. Screening End Code 3	0.75	0.68	0.87	0.39	0.66
Ageband4 vs. Screening End Code 4	0.67	0.77	0.86	0.39	0.66
Screening variation2 vs. Ageband2	0.00	0.00	0.81	0.41	0.63
Screening variation1 vs. Round length1	0.00	0.00	0.59	0.16	0.51
Screening variation2 vs. Round length2	0.01	0.03	1.00	0.33	0.71
Screening variation4 vs. Round length4	0.14	0.13	0.51	0.23	0.45
Screening End Code 1 vs. Attendance2	0.01	0.01	0.41	0.29	0.38
Screening End Code 2 vs. Attendance3	0.44	0.21	0.54	0.38	0.47
Screening End Code 3 vs. Attendance4	0.23	0.24	0.44	0.31	0.40
Attendance 1 vs. 2	0.14	0.11	0.41	0.41	0.38
Attendance 2 vs. 3	0.28	0.29	0.63	0.44	0.53
Attendance 3 vs. 4	0.24	0.33	0.49	0.35	0.44
Postal area vs. Screening variation1	0.00	0.01	0.20	0.05	0.20
Postal area vs. Screening variation2	0.00	0.01	0.21	0.06	0.20
Postal area vs. Screening variation3	0.00	0.01	0.20	0.05	0.20
Postal area vs. screening variation4	0.00	0.02	0.40	0.10	0.40

III. RESULTS AND DISCUSSION

A. Association coefficients

It has been found that the age band of the woman at invitation is not associated with a cancer result or from being involved in a false positive outcome.

All other pairs have certain levels of dependency and association (Table II).

Particularly high associations have been found between the age band of the women, their attendance to the screening episode and the respective screening end code, (except for the first episode where association is not observed).

In addition, a high association has been observed between the age band and the screening variation, particularly for the 2nd episode.

Screening variation, as expected, has a big impact on the round length of the screening episode. Furthermore, screening variation in the previous episode is associated with screening variation in the following, but this association is not very strong. On the other hand, attendance to previous episodes is associated with attendance to the following one.

A medium level of association is observed between the screening end code to the previous episode and future attendance. There is also a very small association with the screening variation and the screening end code in the previous episode, but the association is not pronounced.

The postal area in which the women live at invitation has small association with the occurrence of screening variation but the association with the attendance is very small.

Also, a very small association between the Townsend deprivation score, the attendance, screening variation and screening end code of the women for each episode has been observed.

Diagrams of the intra- and inter-screening episode associations are shown in Figs. 1 & 2.

B. Predictors

A very high prediction can be obtained for the attendance given the age band of the women in a given episode. In particular, if analysed for each particular episode it can be observed that, for the second one, the highest probabilities of non attendance are for women younger than 50 (26%) or older than 64 (22%). For the third episode, non-attendance to invitation assumes high probabilities in those women younger than 55 at invitation (34% if in band 50-54 and 100% if younger than 50). On the other hand, in the fourth episode, the non-attendance probability reaches its maxima in those women with age bands within 50-64 at invitation (24% each). As a conclusion, for the prediction of attendance given the age band, it can be said that each episode has its own characteristics. In order to improve attendance, different age groups of women should be targeted taking into account the screening episode in which they are invited.

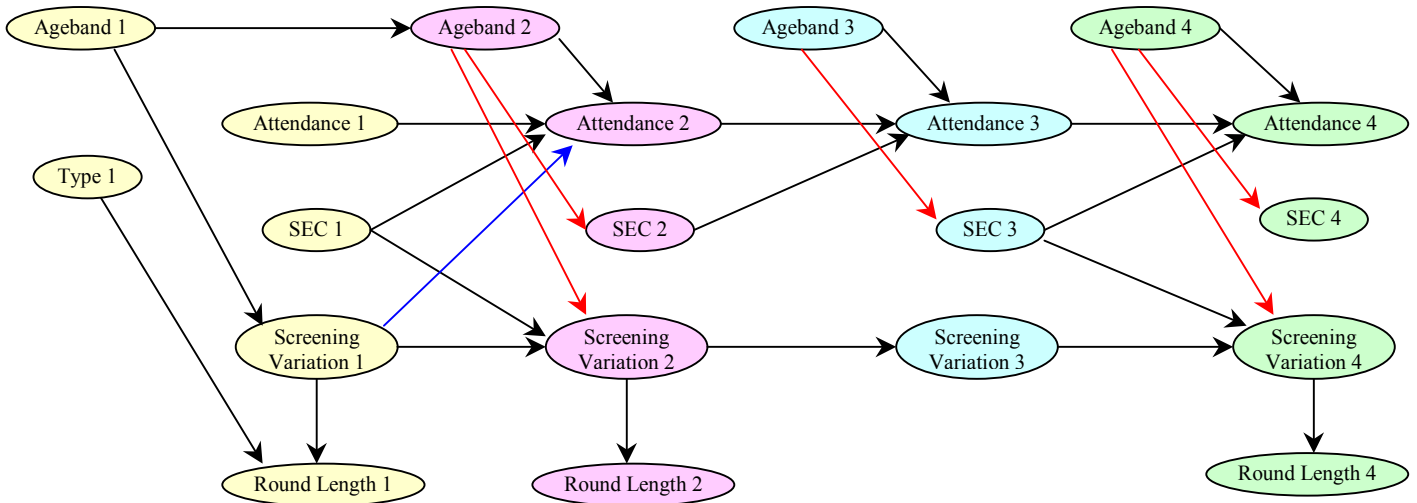


Fig 1. Association flow in intra- and inter-screening episode.

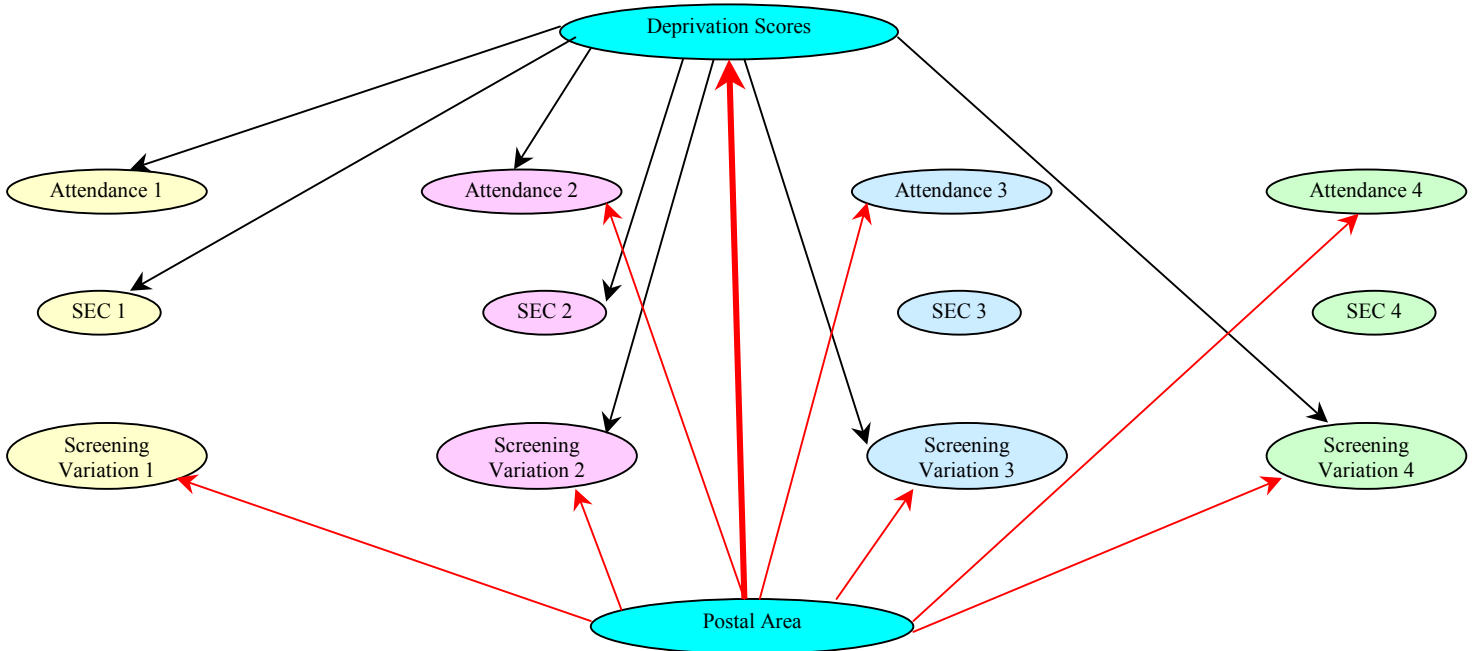


Fig. 2. Townsend deprivation scores and Postal area associations by episodes.

The age band is also an accurate predictor of the screening end code for the episode (except for the first one). Analysing each episode individually, in the second episode, there exists a preponderance of only S- or S+A- end codes. Women under 50 years old have the lowest probabilities of an S- result. The women with higher probabilities of having a cancer result are those who are older than 64 or younger than 50. However, this latter group is more at risk of having a positive or abnormal screening which may not end up in cancer (false positives). In the third episode, the most expected outcomes are S- or S+A- independently of the age group. Only those aged 55-64 have some risk (but very small) of having a cancer outcome for this screening episode. For the fourth

episode, the most probable screening outcome is S-, independently of the age band. For women aged 55-64 there is a small probability (1%) of having an S+A- outcome. Also, for this age group there is a very small risk (<1%) of having a cancer detected. It can be concluded that, in general, over all the episodes, the highest probability is for a screening end code of S- to occur. A result of S+A- frequently occurs in the 2nd and 3rd episodes but decreases considerably in the 4th episode. For the 3rd and 4th episodes the women with more risk of having a cancer detected are those aged between 55-64, whereas for the 2nd episode the women with more risk are those younger than 50 or older than 64.

TABLE III
PREDICTION OF ATTENDANCE GIVEN THE SCREENING END CODE (SEC) OF THE PREVIOUS EPISODE

1 st Episode SEC	Probability of Attendance 2 nd Episode	2 nd Episode SEC	Probability of Attendance 3 rd Episode	3 rd Episode SEC	Probability of Attendance 4 th Episode
No cancer detected	86%	No cancer detected	86%	No cancer detected	92%
Cancer detected	16%	Cancer detected	15%	Cancer detected	9%
S+Aabn H+	50%	S+A+CabnH-	0%	S+AabnC+W+H+	100%
S+A+H+	50%	S+A+C+H-	0%	S+A+H+	50%
SabnAabnC+H+	50%	S+A+C-	0%		
		SabnA+H+	100%		
		SabnAabnC+H+	100%		

Although not as equally significant as age band, prediction of attendance given the screening end code of the previous episode can nevertheless be achieved. Analysis by episode is shown in Table III.

As a conclusion in general, independently of the episode, if the screening end code for the previous episode is negative, then there is high probability that the women will attend the following episode. However, if the screening end code is positive, then the probability is high for non-attendance. This later result was expected given the fact that these women go onto regular symptomatic follow up with annual mammography.

Similarly, prediction for attendance to the episode given the attendance pattern to the previous one is possible. In particular, if the women attend in response to the invitation in the previous episode, they are more likely to attend at the next one too. However, if they did not attend at the previous episode, then there are high probabilities that they will not respond to the invitation for the following episode either [2].

IV. CONCLUSIONS

Numerous conclusions can be drawn from the present study. Amongst the most important, the following can be mentioned,

- In order to improve attendance to the screening episodes, different age groups of women should be targeted depending on the episode of invitation they are in.
- The most probable screening end code are S- and S+A-, but the risk of cancer by age group changes with the episodes. For the 1st and 2nd episodes these include

women younger than 50 or older than 64, whereas for the 3rd and 4th episodes those aged 55-64 are included.

- Women with negative screening end code have high probabilities of attendance but, for those with positive end code, the opposite applies.
- The higher the Townsend deprivation score, the smaller the probabilities of attendance of women at invitations. Particularly low are the probabilities of attendance for those women living in areas with Townsend score 5 or higher (more deprived).
- Attendance to a previous episode is a good predictor of attendance to the following. The opposite also applies.

REFERENCES

- [1] H.E. Arochena, A. Whitford, R.N.G. Naguib, M.E. Wheaton and M.G. Wallis, "Descriptive Analysis of Parameters Affecting a UK Breast Cancer Screening Unit," *Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, Chicago, USA, July 2000 (on CD).
- [2] H.E. Arochena, A. Whitford, R.N.G. Naguib, H. Mashhoudy, M.E. Wheaton and M.G. Wallis, "Analysis of Attendance and Screening Variation in a UK Breast Cancer Screening Unit," *Proceedings 2000 IEEE EMBS International Conference on Information Technology Applications in Biomedicine ITAB 2000*, Virginia, USA, November 2000, pp.332-337.
- [3] H.E. Arochena, R.N.G. Naguib, H. Mashhoudy, M.E. Wheaton and M.G. Wallis, "Screening Variation: Its Impact on Uptake and Coverage for the Breast Screening Programme," *Proceedings European Conference on Cancer Strategies & Outcomes*, Edinburgh, United Kingdom, March 2001, European Journal Of Cancer Vol. 37, Supplement 2, March 2001, pp. S110.
- [4] P. Townsend, P. Phillimore and A. Beattie, *Health and deprivation: inequality and the north*. London: Croom Helm, 1988.
- [5] SPSS for Windows 9.0.0, *On-line help*, SPSS Inc., 1989-1999.
- [6] R.A. Cooper and A.J. Weeks, *Data, Models and Statistical Analysis*. Philip Allan Publishers Limited, 1983.
- [7] B. Everitt, *The Analysis of Contingency Tables*. Chapman and Hall, 1977.